Defence Research and Development Canada | Recherche et developpement pour la defense Canada

DEFENCE R&D DÉFENSE

Ionizing Radiation Detection Meter Systems Usability evaluation

Y. Severs
R. Hughes
D. Tack
L.L.M. Bossi

Terms of release: This document contains proprietary information. It is provided to the recipient on the understanding that proprietary and patent rights will be respected

Defence R&D Canada – Toronto
Technical Memorandum
DRDC Toronto TM 2002-208
December 2002

Canada

Ionizing Radiation Detection Meter Systems usability evaluation

Y Severs

R Hughes

D. Tack

L. L M. Bossi

Terms of release This document contains proprietary information. It is provided to the recipient on the understanding that proprietary and patent rights will be respected

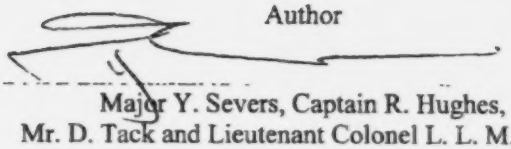
Defence R&D Canada – Toronto

Technical Memorandum

DRDC Toronto TM 2002-208

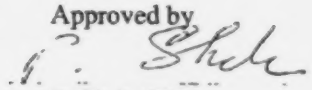
December 2002

Author



Major Y. Severs, Captain R. Hughes,
Mr. D. Tack and Lieutenant Colonel L. L. M. Bossi

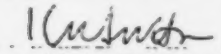
Approved by



P. Shek, PhD

Head, Operational Medicine Section

Approved for release by



K. M. Sutton

Chair, Document Review and Library Committee

© Her Majesty the Queen as represented by the Minister of National Defence, 2002

© Sa majesté la reine, représentée par le ministre de la Défense nationale, 2002

Abstract

In support of the Canadian Forces Nuclear Detection, Identification and Dosimetry project a usability trial of three Ionizing Radiation Detection Meter Systems (IRDMS) was conducted 18-22 June 2001 at Connaught Ranges. The systems evaluated: System 1: ANPDR 77 Operational System with PDR 77 Alpha/Beta probes and GPS (Global Positioning System) Training System; System 2: ADM 300C Operational System with ABP100 Alpha/Beta probes and GPS Training System, and System 3: RAM R200 Operational System with PA 100M and RG12 Alpha/Beta probes and GPS Training System. Twelve military personnel (Private to Captain) determined compliance with the bid evaluation criteria. The counter-balanced study, based on a Latin square design, was divided into two assessment phases, an operational phase and a training system phase, which rotated three groups (of four participants each) concurrently through each IRDMS. Groups were trained in set-up, maintenance and operation of each bid IRDMS prior to evaluation. The operational evaluation consisted of a test/maintenance stand, an indoor/outdoor operational stand, and an indoor/outdoor training system stand. Participants rated the utility and usability of each system, by measuring ease of use/maintenance under a range of relevant operational and clothing conditions, compatibility with soldier tasks, compatibility with clothing, compatibility with equipment and operational environments, and ease of learning (including an assessment of initial and refresher training implications). A 7-point Likert-like acceptability rating scale, which covered a range from completely unacceptable (1) to completely acceptable (7), was used to evaluate the measures. The combined weighted mean scores were 584.3 for System 1 (70%), 682.0 for System 2 (81%) and 756.9 for System 3 (90%) out of a total possible score of 840. Participants reported mean scores for System 1 between barely unacceptable and borderline for both indoor and outdoor operations. System 2 was better, whereby 11 of the 15 questions resulted in mean scores within the borderline to barely acceptable range. System 3 rated reasonably to completely acceptable on all questions, with the exception of prevention of contact with contaminated surfaces which scored barely acceptable. Under set-up and maintenance criteria little difference was found between the 3 systems, although System 1 rated consistently lower than System 2 which rated consistently lower than System 3. System 1 consistently rated barely acceptable to reasonably acceptable, and System 2 barely acceptable to completely acceptable. System 3 was also consistent, whereby 15 of the 16 criteria ranged between reasonably acceptable to completely acceptable. In the training system ratings, System 1 ranged from borderline to barely acceptable, System 2 ranged from barely acceptable to reasonably acceptable and System 3 ranged from reasonably acceptable to completely acceptable. Focus group discussions indicated that System 1, with an extremely small display and function buttons was too complex, with the potential for creating performance error. Another key problem was that participants could not comfortably operate System 1 and System 2 with the extension probe for 20 minutes, a realistic requirement in an operational setting. System 3 was found to be comfortable and balanced. Although some minor changes were suggested in focus group discussions System 3 was found to be the most acceptable system operationally, with an acceptable training system.

Résumé

En soutien au projet de détection, d'identification et de dosimétrie nucléaires des Forces canadiennes, nous avons effectué au polygone de Connaught, entre les 18 et 22 juin 2001, des essais sur la facilité d'emploi de trois systèmes de détection et de mesure de la radioactivité. Nous avons évalué les trois systèmes suivants : (1) un appareil opérationnel AN-PDR 77 avec des sondes alpha/bêta PDR 77 et un système GPS (Système mondial de localisation) pour l'entraînement; (2) un appareil opérationnel ADM 300C avec des sondes alpha/bêta ABP100 et un système GPS pour l'entraînement; (3) un appareil opérationnel RAM R200 avec des sondes alpha/bêta PA 100M et RG12 et un système GPS pour l'entraînement. Douze membres des forces (des grades de soldat à capitaine) ont mesuré la conformité des systèmes avec les critères d'évaluation des soumissions. L'étude équilibrée, basée sur le carré latin, comportait deux phases d'évaluation : une phase opérationnelle et une autre d'entraînement à l'utilisation du système. Trois groupes de quatre participants ont utilisé chaque système en succession. On a formé les groupes au réglage, à l'entretien et à l'utilisation de chaque système avant qu'ils les évaluent. L'évaluation opérationnelle comportait un banc d'essai et d'entretien, un autre pour évaluer le fonctionnement à l'intérieur et à l'extérieur et un dernier banc pour la formation à l'intérieur ou à l'extérieur. Les participants ont évalué l'utilité et la facilité d'emploi de chaque système en évaluant s'ils étaient faciles à utiliser et à entretenir pour une gamme de conditions opérationnelles et de tenues militaires; sa compatibilité avec les tâches des soldats, la tenue portée, l'équipement et l'environnement des opérations; ainsi que la facilité d'apprentissage (notamment l'évaluation de l'effet des formations initiales et d'appoint). Pour évaluer les mesures, nous avons utilisé une échelle de notation de sept points, analogue à celle de Likert, s'étendant de « complètement inacceptable » (note 1) à « complètement acceptable » (note 7). Nous avons obtenu des cotes combinées des moyennes pondérées de 584,3/840 (70 %) pour le système 1, de 682,0/840 (81 %) pour le système 2 et de 756,9/840 (90 %) pour le système 3. Les participants ont attribué au système 1 des notes moyennes entre « à peine inacceptable » et « médiocre » pour son utilisation à l'intérieur ou à l'extérieur. Ils ont mieux aimé le système 2 et lui ont attribué une note moyenne allant de « médiocre » à « à peine acceptable », à 11 questions sur 15. Les participants ont accordé au système 3 des notes allant de « raisonnablement acceptable » à « complètement acceptable » pour toutes les questions, hormis celle sur la prévention du contact avec les surfaces contaminées à laquelle ils ont répondu : « à peine acceptable ». Les différences entre les trois systèmes quant aux critères du réglage étaient minimes. Cependant, les notes du système 1 étaient toutes plus basses que celles du système 2, lesquelles étaient toutes plus basses que celles du système 3. Les participants ont généralement attribué au système 1 une note allant de « à peine acceptable » à « raisonnablement acceptable » et au système 2, de « à peine acceptable » à « complètement acceptable ». Les notes attribuées au système 3 étaient cohérentes : à 15 questions sur les 16, les participants ont attribué une note allant de « raisonnablement satisfaisant » à « complètement satisfaisant ». Au chapitre des notes sur l'entraînement, on a donné au système 1 une note allant de « médiocre » à « à peine acceptable », au système 2 de « à peine acceptable » à « raisonnablement acceptable » et au système 3 de « raisonnablement acceptable » à « complètement acceptable ». Le groupe de discussion a indiqué que l'affichage trop petit et les boutons de fonction trop complexes du système 1 pouvaient provoquer des erreurs de fonctionnement. Les essais ont révélé un autre problème important : les participants ont indiqué qu'ils ne pouvaient pas utiliser aisément les systèmes 1 et 2 avec la sonde pendant 20 minutes, une exigence prévisible dans un contexte

opérationnel. On a trouvé que le système 3 était d'un emploi facile et était mécaniquement équilibré. Le groupe de discussion a suggéré des modifications mineures au système 3, lequel a été considéré comme le système le plus acceptable pour les opérations, après un entraînement adéquat.

This page intentionally left blank.

Executive summary

In support of the Canadian Forces Nuclear Detection, Identification and Dosimetry project a usability trial of three Ionizing Radiation Detection Meter Systems (IRDMS) was conducted 18-22 June 2001 at Connaught Ranges. The systems evaluated were: System 1: ANPDR 77 Operational System with PDR 77 Alpha/Beta probes and GPS (Global Positioning System) Training System; System 2: ADM 300C Operational System with ABP100 Alpha/Beta probes and GPS Training System; and System 3: RAM R200 Operational System with PA 100M and RG12 Alpha/Beta probes and GPS Training System. Twelve military personnel determined compliance with the bid evaluation criteria. The counter-balanced study, based on a Latin square design, was divided into two assessment phases, an operational phase and a training system phase, which rotated three groups (of four participants each) concurrently through each IRDMS. Groups were trained in set-up, maintenance and operation of each bid IRDMS prior to evaluation. The operational evaluation consisted of a test/maintenance stand, an indoor/outdoor operational stand, and an indoor/outdoor training system stand. Participants rated the utility and usability of each system, by measuring ease of use/maintenance under a range of relevant operational and clothing conditions, compatibility with soldier tasks, compatibility with clothing, compatibility with equipment and operational environments, and ease of learning (including an assessment of initial and refresher training implications). A 7-point acceptability rating scale which covered a range from completely unacceptable (1) to completely acceptable (7) was used to evaluate the measures. The combined weighted mean scores were 584.3 for System 1 (70%), 682.0 for System 2 (81%) and 756.9 for System 3 (90%) out of a total possible score of 840. Participants reported mean scores for System 1 between barely unacceptable and borderline for both indoor and outdoor operations. System 2 was better, whereby 11 of the 15 questions resulted in mean scores within the borderline to barely acceptable range. System 3 rated reasonably to completely acceptable on all questions, with the exception of prevention of contact with contaminated surfaces which scored barely acceptable. Under set-up and maintenance criteria little difference was found between the 3 systems, although System 1 rated consistently lower than System 2 which rated consistently lower than System 3. System 1 consistently rated barely acceptable to reasonably acceptable, and System 2 barely acceptable to completely acceptable. System 3 was also consistent, whereby 15 of the 16 criteria ranged between reasonably acceptable to completely acceptable. In the training system ratings, System 1 ranged from borderline to barely acceptable, System 2 ranged from barely acceptable to reasonably acceptable and System 3 ranged from reasonably acceptable to completely acceptable. Focus group discussions indicated that System 1, with an extremely small display and function buttons was too complex, with the potential for creating performance error. Another key problem was that participants could not comfortably operate System 1 and System 2 with the extension probe for 20 minutes, a realistic requirement in an operational setting. System 3 was found to be comfortable and balanced. Although some minor changes were suggested in focus group discussions System 3 was found to be the most acceptable system operationally, with an acceptable training system.

Severs, Y., Hughes, R., Tack, D., Bossi, L. 2002. Ionizing Radiation Detection Meter Systems usability evaluation. DRDC Toronto TM 2002-208 Defence R&D Canada – Toronto.

Sommaire

En soutien au projet de détection, d'identification et de dosimétrie nucléaires des Forces canadiennes, nous avons effectué au polygone de Connaught, entre les 18 et 22 juin 2001, des essais sur la facilité d'emploi de trois systèmes de détection et de mesure de la radioactivité. Nous avons évalué les trois systèmes suivants : (1) un appareil opérationnel AN-PDR 77 avec des sondes alpha/bêta PDR 77 et un système GPS (Système mondial de localisation) pour l'entraînement, (2) un appareil opérationnel ADM 300C avec des sondes alpha/bêta ABP100 et un système GPS pour l'entraînement; (3) un appareil opérationnel RAM R200 avec des sondes alpha/bêta PA 100M et RG12 et un système GPS pour l'entraînement. Douze membres des forces ont mesuré la conformité des systèmes avec les critères d'évaluation des soumissions. L'étude équilibrée, basée sur le carré latin, comportait deux phases d'évaluation : une phase opérationnelle et une autre d'entraînement à l'utilisation du système. Trois groupes de quatre participants ont utilisé chaque système en succession. On a formé les groupes au réglage, à l'entretien et à l'utilisation de chaque système avant qu'ils les évaluent. L'évaluation opérationnelle comportait un banc d'essai et d'entretien, un autre pour évaluer le fonctionnement à l'intérieur et à l'extérieur et un dernier banc pour la formation à l'intérieur ou à l'extérieur. Les participants ont évalué l'utilité et la facilité d'emploi de chaque système en évaluant s'ils étaient faciles à utiliser et à entretenir pour une gamme de conditions opérationnelles et de tenues militaires; sa compatibilité avec les tâches des soldats, la tenue portée, l'équipement et l'environnement des opérations; ainsi que la facilité d'apprentissage (notamment l'évaluation de l'effet des formations initiales et d'appoint). Pour évaluer les mesures, nous avons utilisé une échelle de notation de sept points s'étendant de « complètement inacceptable » (note 1) à « complètement acceptable » (note 7). Nous avons obtenu des cotes combinées des moyennes pondérées de 584,3/840 (70 %) pour le système 1, de 682,0/840 (81 %) pour le système 2 et de 756,9/840 (90 %) pour le système 3. Les participants ont attribué au système 1 des notes moyennes entre « à peine inacceptable » et « médiocre » pour son utilisation à l'intérieur ou à l'extérieur. Ils ont mieux aimé le système 2 et lui ont attribué une note moyenne allant de « médiocre » à « à peine acceptable », dans 11 questions sur 15. Les participants ont accordé au système 3 des notes allant de « raisonnablement acceptable » à « complètement acceptable » pour toutes les questions, hormis celle sur la prévention du contact avec les surfaces contaminées à laquelle ils ont répondu : « à peine acceptable ». Les différences entre les trois systèmes quant aux critères du réglage étaient minimes. Cependant, les notes du système 1 étaient toutes plus basses que celles du système 2, lesquelles étaient toutes plus basses que celles du système 3. Les participants ont généralement attribué au système 1 une note allant de « à peine acceptable » à « raisonnablement acceptable » et au système 2, de « à peine acceptable » à « complètement acceptable ». Les notes attribuées au système 3 étaient cohérentes : à 15 questions sur les 16, les participants ont attribué une note allant de « raisonnablement satisfaisant » à « complètement satisfaisant ». Au chapitre des notes sur l'entraînement, on a donné au système 1 une note allant de « médiocre » à « à peine acceptable », au système 2 de « à peine acceptable » à « raisonnablement acceptable » et au système 3 de « raisonnablement acceptable » à « complètement acceptable ». Le groupe de discussion a indiqué que l'affichage trop petit et les boutons de fonction trop complexes du système 1 pouvaient provoquer des erreurs de fonctionnement. Les essais ont révélé un autre problème important : les participants ont indiqué qu'ils ne pouvaient pas utiliser aisément les systèmes 1 et 2 avec

la sonde pendant 20 minutes, une exigence prévisible dans un contexte opérationnel. On a trouvé que le système 3 était d'un emploi facile et était mécaniquement équilibré. Le groupe de discussion a suggéré des modifications mineures au système 3, lequel a été considéré comme le système le plus acceptable pour les opérations, après un entraînement adéquat.

Severs, Y., Hughes, R., Tack, D., Bossi, L. 2002. Ionizing Radiation Detection Meter Systems usability evaluation. DRDC Toronto TM 2002-208 Defence R&D Canada – Toronto.

This page intentionally left blank.

Table of contents

Abstract i

Résumé ii

Executive summary v

Sommaire vi

Table of contents ix

List of figures xi

List of tables xi

Acknowledgements xii

Introduction 1

- Background 1
- Aim 1

Method 3

- Overview 3
- Participants 3
- System Conditions 3
- Protocol 8
- Data Collection 10
- Statistics 10

Results and Discussion 12

- Results 12
- Discussion 15
- Operations 16
- Set-up and Maintenance 17
- Training 18
- Ratings and Summary 20

Conclusions 21
Compliance with Request for Proposal and Improvements 21
References 23
Annex A IRDMS Usability Evaluation Questionnaires 24
List of Abbreviations 30

List of figures

Figure 1. System 1, ANPDR 77 (extension) 4

Figure 2. System 1, ANPDR 77 (meter and probes) 4

Figure 3. System 1 Training System, ANPDR 77 5

Figure 4. System 2, ADM 300C (extension and probe) 5

Figure 5. System 2, ADM 300C (meter) 6

Figure 6. System 2 Training System, ADM 300C 6

Figure 7. System 3, RAM R 200 7

Figure 8. System 3 Training System, RAM R 200 7

Figure 9. Seven-point acceptability rating scale 10

List of tables

Table 1. Schedule and Rotation 8

Table 2. Operational Results 12

Table 3. Set-up and Maintenance Results 13

Table 4. Training Results 14

Acknowledgements

The authors wish to acknowledge the assistance of Mr. Donovan, Mr. Williams, Mr. Turbide and CWO MacDonald for their assistance during the Ionizing Radiation Detection Meter Systems Usability Evaluation. We also wish to extend our thanks to the military volunteers, without their professionalism, candid observations and worthy suggestions the success of this trial would not be possible.

Introduction

Background

As part of capital Project G2199, the Canadian Forces Nuclear Detection, Identification and Dosimetry Project, a usability evaluation was performed for the bid evaluation of various candidate Ionizing Radiation Detection Meter Systems (IRDMS). This project will replace the in-service ionizing radiation detection meter (ADM 300) which is out-dated from a technological standpoint and not ideally suited for manual field surveys or usable for personnel without special training in radiation monitoring. The capital project will provide modern nuclear detection, identification and dosimetry equipment systems and the associated IRDMS educational and training tools needed to develop and maintain competence at the individual, unit and formation levels.

IRDMS will provide the capability to detect low levels of ionizing radiation and prevent accidental exposures of others to ionizing radiation. It will also provide the Canadian Forces with the capability to manage essential exposure levels in accordance with the Nuclear Safety and Control Act (NSCA), Bill C-23 and the internationally accepted principle of keeping the exposures as low as reasonably achievable.

The System will be used to conduct operations in areas that may be contaminated by low level radiation or tactical level radiation and as such will be used by a cross-section of military operational and support occupations in the Canadian Forces. Individuals with little or no understanding of nuclear technologies will be required to use the IRDMS for reconnaissance, survey, monitoring and both wide- and small-area detection training. This could include military personnel ranging in rank from Corporal to Captain with variable Radiation Safety training and experience (from untrained to qualified Radiation Safety Officers (RadSO)).

IRDMS will also be used to carry out reconnaissance of potentially contaminated areas in order to provide commanders with early warning of contaminated operating areas. Thus, it is intended that the IRDMS be compatible with either vehicle or foot borne reconnaissance and used to detect alpha, beta and gamma radiation hazards in order to establish gamma radiation dose rate contours, or the extent of alpha or beta contamination. The system will also be used to carry out detailed surveys and monitoring of potentially contaminated personnel and equipment for alpha, beta, and gamma radiation hazards.

Three IRDMS candidate bids were received which were comprised of two components, an Operational System for use in real situations and a Training System intended to represent operational equipment thus allowing training without exposing individuals to radioactive sources.

Aim

The aim of the usability evaluation was to determine the compliance of IRDMS plus accompanying training systems, with the usability aspects of the bid evaluation criteria used

by the acquisition project G2199, based upon the criteria specified in the Request for Standing Offer (RFSO).

Method

Overview

The usability evaluation was conducted at the Connaught Ranges in Ottawa from 18 to 22 June 2001. Twelve personnel from different military occupations, of varying rank and ranging in experience participated in the trial. Each participant received familiarization training on each of the candidate systems prior to performing a range of set-up, maintenance and operationally relevant tasks using each system. Evaluations of each system were collected via questionnaires and focus group interviews and then collated to identify the ratings of each system.

Participants

The twelve participants involved in the usability evaluation were drawn from various units throughout the CF. Their military occupations were: two Aerospace Control Operators (168), one Flight Engineer (091), one Maritime Engineer (44B), one Marine Engineer Technician (313), one Infantry soldier (031), one Field Engineer (041), one Maritime Surface and Subsurface Technician (71B), two Medical Assistants (711), and two Supply Technicians (911). The subject's ranks ranged from Private (1), to junior non commissioned member (NCM) (3), to senior NCM (6) and to officer (2). The participants exposure with similar equipment varied; two were limited with only yearly refresher training, four were average where equipment was used periodically during the year, and seven were considered advanced where similar equipment was used in their daily performance of duties.

System Conditions

1. The three systems were examined in this usability evaluation are as follows:

    a. System 1, ANPDR 77, provided by APTEC-NRC included IRDMS Operations Kit with PDR 77 Alpha and Beta probes and GPS (Global Positioning System) training system;

    b. System 2, ADM 300C, provided by APTEC-NRC included GPS Module (ADM 300C), Alpha Beta probes (ABP 100) and GPS training system; and

    c. System 3, RAM R 200, provided by MGP Instruments included and Operational System with Alpha (PA 100M) and Beta (RG12) probes and GPS training system (System 3's training system was a prototype and not configured to resemble the operational system).

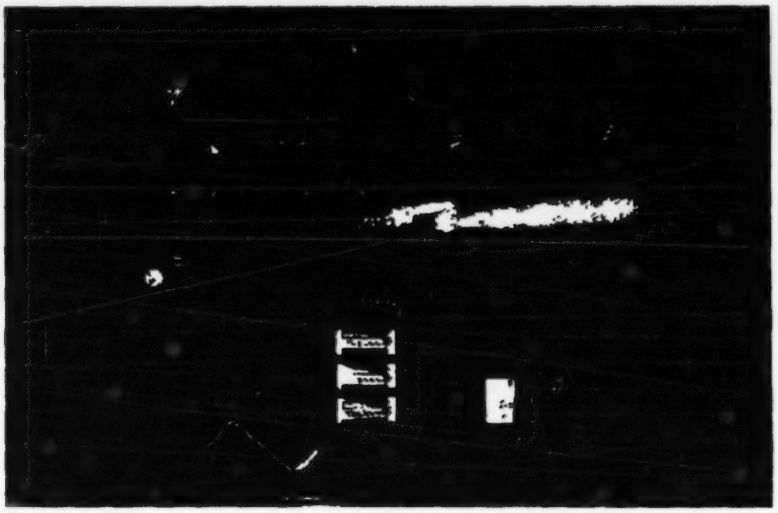


Figure 1. *System 1, ANPDR 77 (extension)*

Figure 2. *System 1, ANPDR 77 (meter and probes)*

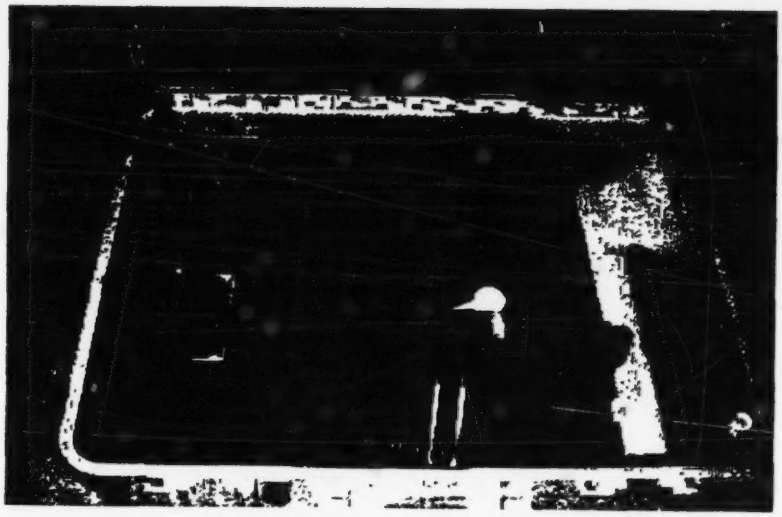


Figure 3. *System 1 Training System, ANPDR 77*

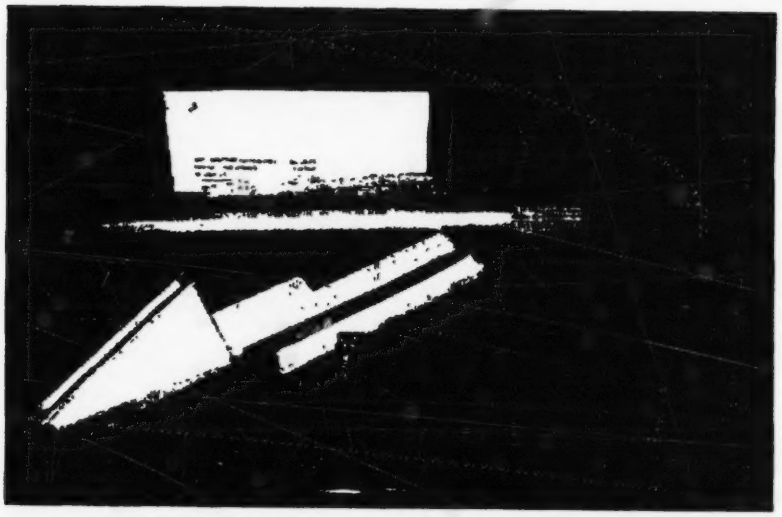


Figure 4. *System 2, ADM 300C (extension and probe)*

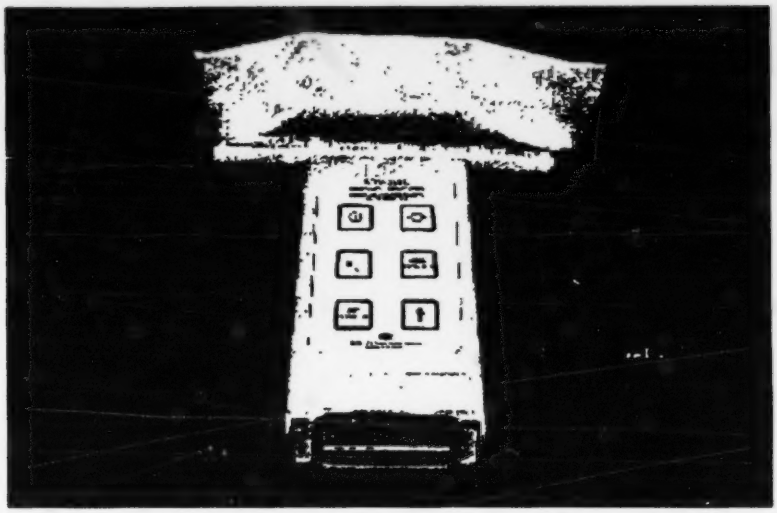


Figure 5. System 2, ADM 300C (meter)

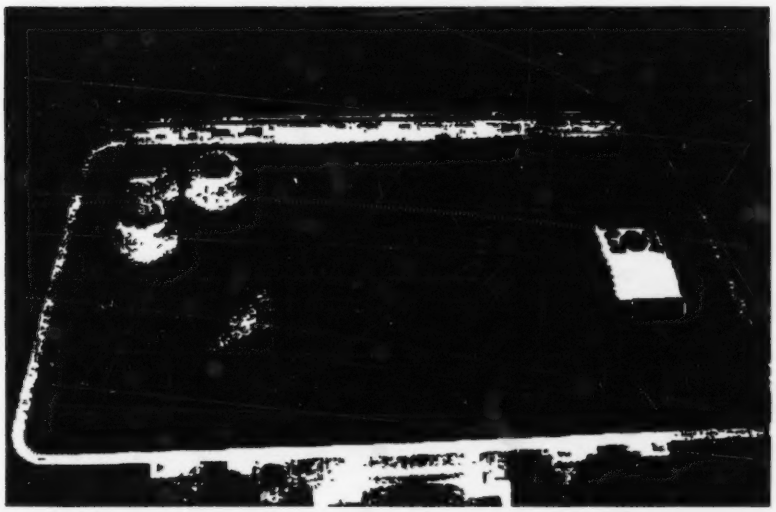


Figure 6. System 2 Training System, ADM 300C

***Figure 7.** System 3, RAM R 200*

***Figure 8.** System 3 Training System, RAM R 200*

Protocol

The experiment was conducted in two phases, an operational evaluation phase and a training evaluation phase. In both phases participants were divided into three groups of four. Group assignments were balanced to ensure that there was an even distribution of NBC (Nuclear, Biological and Chemical) and/or radiation training and skills throughout all groups. Therefore, group selection was based on prior NBC and radiation training, individual NBC skills of each participant, military experience, rank, as well as, the military element of the participants. The trial design had both the IRDMS systems and test stands for each phase assigned to groups in a counter-balanced fashion using a Latin square design. This balanced repeated block design was used to ensure there was no bias due to order of presentation, learning, or subject experience. Table 1 details the trial schedule and task rotation.

Table 1. Schedule and Rotation

DAY 1	DAY 2			DAY 3			DAY 4		
	Group 1	***Group 2***	***Group 3***	***Group 1***	***Group 2***	***Group 3***	***Group 1***	***Group 2***	***Group 3***
Welcome Bnef	System 1 Refresher Trg	System 2 Refresher Trg	System 3 Refresher Trg	System 2 Refresher Trg	System 3 Refresher Trg	System 1 Refresher Trg	System 1 Trg Indoor Stand	System 2 Trg Indoor Stand	System 3 Trg Outdoor Stand
Training on System 1	Ops Indoor Stand	Ops Outdoor Stand	Set-up/ Maint Stand	Set-up/ Maint Stand	Ops Indoor Stand	Ops Outdoor Stand	System 2 Trg Outdoor Stand	System 1 Trg Indoor Stand	System 2 Trg Indoor Stand
	Ops Outdoor Stand	Set-up/ Maint Stand	Ops Indoor Stand	Ops Indoor Stand	Ops Outdoor Stand	Set-up/ Maint Stand	System 3 Trg Indoor Stand	System 3 Trg Outdoor Stand	System 1 Trg Indoor Stand
Lunch									
	Refresher Trg	Refresher Trg	Refresher Trg		Operational Phase			Training Phase	
Training on System 2	Ops Indoor Stand	Ops Outdoor Stand	Set-up/ Maint Stand	Focus	Group	Discussion	Focus	Group	Discussion
Training on System 3	Ops Outdoor Stand	Set-up/ Maint Stand	Ops Indoor Stand	Training	On Systems	Training	Total	Acceptability Of System	
	Set-up/ Maint Stand	Ops Indoor Stand	Ops Outdoor Stand						

On the first day participants were trained to use the components of each system, and prior to each operational and training evaluation each group completed refresher training. In the evaluation, tasks were divided into three phases: set-up/maintenance, operations and training. The phases were derived directly from the specific requirements identified in the Statement of Work, providing the participants an opportunity to acquaint themselves with the three systems and providing a basis upon which to evaluate the system. As representative of an operational setting, participants worked in teams of two while using each system.

The set-up/maintenance evaluation stand allowed participants to conduct various tasks associated with assembling the systems and confirming their working condition. Tasks included: assembling each IRDMS; confirming operational serviceability; setting auditory and visual alarms for various dose levels; and, conducting basic maintenance on each system. During the operational evaluation participants evaluated each operational portion of the IRDMS in both indoor and outdoor reconnaissance/survey conditions to assess acceptability under the various conditions representative of potential contamination. In the training system evaluation, the alpha/beta portion of the systems were evaluated using the indoor and outdoor operation reconnaissance/survey protocol, as the training systems are intended to resemble each of the three operational IRDMS. For the gamma portion of the IRDMS training systems, which utilised GPS to simulate gamma contamination on a computer, participants were trained in the classroom setting, and asked to evaluate the process of setting up a computer gamma survey. As both Training System 1 and System 2 utilised the same computer GPS based system for gamma simulation, they were taught at the same time. Actual simulation of the GPS gamma training systems was not possible during this trial because a satellite connection could not be established for System 1 and 2, and System 3's GPS was not functional during the trial.

During the Operational evaluation tasks, the radiation check sources used were representative of alpha and beta radiation. These sources are normally used in radiation training and posed no undue risk to subjects. However, as an added precaution, only project staff handled the check sources to reduce unnecessary exposure. Also, as the radiation check sources were positioned in new locations by project staff prior to group rotation a learning effect across system tests was avoided thus keeping the tasks operationally realistic

In the Training evaluations, the three systems used a variety of non-radioactive substances to represent alpha and beta check sources in order to elicit and simulate positive responses. System 1's training system utilised tin foil to represent alpha and beta contamination. Both System 2 and System 3 used liquid and powder stimulants as check sources, and additionally System 3 used normal magnets to represent alpha and beta contamination. The manufacturers did not identify the chemical stimulate names and accordingly they were referred to as yellow stimulant liquid, and grey, blue or green stimulant powders. The liquid and powder stimulants were considered non-toxic, disposable as part of household waste and therefore were considered to pose no risk to subjects or staff.

The operational and training evaluation tasks were designed to represent the realm of operational conditions under which the systems would be used. This included lighting conditions which ranged from normal to artificial and night conditions. The normal light condition utilised ambient sunlight and overhead fluorescent lighting; the fluorescent light condition utilised only overhead fluorescent lighting; and the low light/night condition was

simulated by a darkened a classroom. The bright sunlight condition were performed outside the classroom in direct sunlight. During the trial, the days were bright and sunny and the evaluations were completed in the early afternoon.

Data Collection

Upon completion of each test stand (set-up/maintenance; operations and training) evaluations were made using usability questionnaires based upon the requirements set up in the Request For Proposal (RFP) (Annex A). Participants rated the utility and usability of each system by rating ease of use/maintenance under range of relevant operational and clothing conditions, compatibility with soldier tasks, compatibility with clothing, compatibility with equipment and operational environments, and ease of learning (including an assessment of initial and refresher training implications). The questionnaire used a Likert-like scale, a commonly used scale in usability trials, with a seven-point acceptability rating scale (Figure 7).

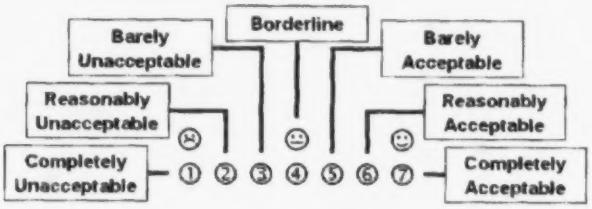


Figure 9. *Seven-point acceptability rating scale*

Upon completion of the study it was determined that certain criteria within the studies' specifications could not be fairly evaluated due to experimental conditions and disparity between systems. Specifications that were not common to all systems (i.e. earphones) could not be compared thus the questions were removed and not tallied in the mean rating and combined weighted tally. These questions are detailed in the result's section.

Focus groups were held after both the operations and training evaluations to better understand participant questionnaire ratings, provide feedback for bidders, permit project staff to negotiate improvements prior to acquisition and compare and contrast the different systems. During focus groups, to further amplify acceptability participants were asked to rate the acceptability of each system by specific criteria, and to rate the most acceptable overall system, as well as comment on possible improvements for each individual system should it be procured.

Statistics

The trial was designed to assess the acceptability of the systems. Participants used the rating scale to rate what was not acceptable (completely unacceptable to less than borderline) and what was acceptable (borderline to completely acceptable). Mean ratings for each questionnaire item were provided to the project management team, as tasked, so that the

usability aspects would contribute to the bid evaluation scoring system (as laid out in the RFP specification). In describing the mean ratings, results were not rounded and actual figures were used to provide guidance on what was acceptable or unacceptable. Consequently, as an example, a barely acceptable description was used to describe a mean rating between 5 and 5.9. No significance statistical testing was conducted.

Results and Discussion

Results

In the Operational evaluations, the mean indoor and outdoor acceptability ratings were used to generate a combined mean rating. The results from the questionnaires are summarized in Tables 2, 3 and 4 below. Table 2 details the results of the Operational evaluation, Table 3 provides the results from the Set-up and Maintenance portion of the questionnaire and Table 4 details the Training results. The number detailed is the mean rating for each question.

Table 2. *Operational Results*

		MEAN ACCEPTABILITY RATING								
	TEST STAND	INDOORS			OUTDOORS			COMBINED		
OPERATION	***System***	***1***	***2***	***3***	***1***	***2***	***3***	***1***	***2***	***3***
3 3 1 Ease of operating the controls		3 9	6 2	6 3	4 3	6 0	6 2	4 1	6 1	6 3
3 6 8 a) Compactness of the operational probes		4 4	5 3	6 8	4 5	4 6	6 7	4 5	4 9	6 7
3 6 8 b) Weight of the operational probes		4 9	5 1	6 6	4 6	4 3	6 7	4 8	4 7	6 6
3 6 12 Effectiveness of the system for preventing inadvertent contact with contaminated surfaces.		4 3	4 7	5 1	3 5	3.4	5 4	3 9	4 0	5 3
3 7 4 Ease of using the probe extension		NA	NA	NA	3 2	4 9	6 8	3 2	4 9	6 8
3 9 2 Ease of using the system to survey for up to six hours on a 20-minute work/10 minute rest cycle		4 3	4 6	6 4	3 5	3.9	6 4	3 9	4 3	6 4
3 9 1 b) Compactness of system (i e meters, cases, probes, and extension wands)		3 3	4 7	6 3	3 6	4 1	6 3	3 5	4 4	6 3
3 9 1 c) Weight of system (i e. meters, cases, probes, and extension wands).		3 8	4 3	6 3	4 4	4 0	6 6	4 1	4 1	6 5
3 9 3 a) Compatibility of System components with combat clothing (i e combats, combat gloves, and sand/wind/dust goggles)		4 3	5 1	6 4	4 8	4 6	6 6	4 5	4 8	6 5
3 9 3 b) Compatibility of System components with NBC TOPP high (i e NBC IPE, NBC gloves, overboots, and C4 mask)		3 2	4 8	6 2	3 6	3 9	6 1	3 4	4 4	6 1
DURABILITY	***System***	***1***	***2***	***3***	***1***	***2***	***3***	***1***	***2***	***3***
3 6 11 Effectiveness of mechanism for protecting the probe, when in use		4 3	4 8	5 6	4 1	3 3	5 6	4 2	4 0	5 6
3 8 1 a) Durability of storage cases		6 3	6 4	6 8	6 0	6 3	6 7	6 1	6 3	6 7
3 8 1 b) Ease of using storage cases		3 8	6 0	6 3	4 1	6 0	6 3	4 0	6 0	6 3
3 8 5 Ease of distinguishing storage cases.		5 9	5 8	6.6	5 8	5 9	6 6	5 9	5 9	6 6
3 9 1 a) Durability of system (i e meters, cases, probes, and extension wands)		4 9	5 2	6 3	4 8	4 4	6 2	4 8	4 8	6 3
TRAINING SYSTEM Vs OPERATIONAL SYSTEM	***System***	***1***	***2***	***3***	***1***	***2***	***3***	***1***	***2***	***3***

5 4 2 Similarity of information displayed by the training system to the operational system							5 5	6 1	6 1
5 4 3 Similarity of operation of the training system to the operational system							5 1	5 5	6 0

In Table 2, ease of using the probe extension (3.7.4) was not scored for the indoor survey and indicates NA (non-applicable); although this is a criteria RFP rating, participants indicated performance of a small space indoor survey was easier without the extension. Therefore, as the extension was not used during the indoor survey and as this represented operational realism the rating was removed from scoring for indoor operations. Similarly questions evaluating whether the training systems represented the operational systems (5.4.2 and 5.4.3) were not rated separately for indoor and outdoor operations. As the questions were intended to rate the similarity of the training system to the operational system, both in display and operation, participants compared all operational components against the training system upon the completion of the training phase. The questions are in fact a rating of the compatibility of the training system and thus it was not necessary to distinguish between indoor and outdoor operations.

Table 3. *Set-up and Maintenance Results*

		MEAN ACCEPTABILITY RATING		
CONFIRM OPERATION	***System***	***1***	***2***	***3***
3 2 6 a) Ease of comprehending the visual display changes in radioactivity levels		5 6	6 1	6 7
3 2 6 b) Ease of comprehending the auditory display changes in radioactivity levels		5 5	5 9	6 7
3 2 7 a) Readability of visual display in normal daylight conditions		6 3	6 1	6 4
3 2 7 b) Readability of visual display in bright sunlight conditions		6 4	5 3	6 3
3 2 7 c) Readability of visual display in fluorescent lighting conditions		6 5	6 3	6 6
3 2 7 d) Readability of visual display in low light / night conditions		4 4	6 6	6 9
3 2 8 a) Ease of silencing the audible display		5 6	6 8	6 8
3 2 8 b) Speed of silencing the audible display		6 1	6 5	6 8
3 2 9 Ease of operating the volume control		NA	NA	NA
3.2 10 Effectiveness of the earphones in high noise conditions		NA	NA	NA
3.2 11 Ease of subtracting background radiation levels during alpha or beta surveys		NA	NA	NA
3 3 2 Ease of switching between gamma, alpha, and beta radiation readings		5 2	6 2	6 1
SET ALARMS	***System***	***1***	***2***	***3***
3 4 2 a) Effectiveness of the visual alarm		5 1	5 7	6 3
3 4 2 b) Effectiveness of the auditory alarm		5 3	6 5	6 4
3 4 5 a) Ease of deactivating/silencing audible alarm before alarm level is reached.		5.7	5 9	6.1
3 4 5 b) Ease of deactivating/silencing audible alarm after alarm level is reached		5 8	6 3	6 4
3 4 6 Ease of setting the alarm set points		4 8	6 2	5 5
3.4.7 a) Ease of activating the alarms		5 7	6 5	5 9
3 4 7 b) Ease of deactivating the alarms		5 6	6 3	6 0

MAINTENANCE	System	1	2	3
3 10 3 Ease of replacing batteries in the field		5 8	6 0	6 8
3 10 7 a) Effectiveness of auditory "low battery" warning		NA	NA	NA
3 10 7 b) Effectiveness of visual "low battery" warning.		NA	NA	NA

In Table 3 five questions (3.2.9, 3.2.10, 3.2.11, 3.10.7 a, 3.10.7 b) were not scored and indicate NA. Ease of operating the volume control (3.2.9) was removed from tally because not all systems had a volume control. Also, effectiveness of the earphones in high noise conditions (3.2.10) was not tallied as all systems did not have this component. Ease of subtracting background radiation levels during alpha or beta surveys (3.2.11) was also not tallied because participants could not perform the function. Thus, they were unable to fairly answer this question. Finally, effectiveness of auditory "low battery" warning (3.10.7a) and effectiveness of visual "low battery" warning (3.10.7 b) were not scored as the low battery conditions could not be replicated during the trial. As participants could not perform the actions and thus could not fairly answer this question, it was deemed appropriate to remove the questions. In this study all efforts were taken to insure that usability comparisons were made between similar pieces of equipment with similar characteristics. Where this was not possible, as describe above, specific questions were removed to ensure fair and proper comparisons.

Table 4. *Training Results*

		MEAN ACCEPTABILITY RATING		
CONTROLS AND PROBES	**System**	**1**	**2**	**3**
3 3 1 a) Ease of learning the controls		5 4	6 4	6 2
3 3 1 b) Ease of comprehending the control markings		5 4	6 5	6 3
3 6 8 a) Compactness of the training probes		5 6	5 6	6 2
3 6 8 b) Weight of the training probes		5 3	5 1	6 3
TRAINING SYSTEM	**System**	**1**	**2**	**3**
5 2 3 Ease of using the training system		5 6	5 7	5 6
5 4 1 Durability of the training system		5 4	5 8	5 8
5 4 4 1 Effectiveness of training system for training widely dispersed ground contamination by multiple gamma, alpha, or beta emitters		3 3	5 3	5 4
5 4 4 2 Effectiveness of training system for training wide airborne contamination by by multiple gamma, alpha, or beta emitters (e g in a plume)		NA	NA	NA
5 4 4 3 Effectiveness of training system for training changes in contamination according to user-defined winds or radioactive decay rates		NA	NA	NA
5 4 4 4 Effectiveness of training system for training the detection of agents only within a few centimetres of the contamination surface		4 2	6 2	5 5
5 4 4 5 Effectiveness of training system for training probe contamination when contacting the contaminated surface.		4 2	5 8	4 3
5 4 4 6 Effectiveness of training system for training monitoring during and after decontamination of personnel and equipment.		4 6	5 5	5 6

5 4 4 7 Effectiveness of training system for training statistical fluctuations in the rate counts	NA	NA	NA
5 4 4 8 Effectiveness of training system for training alarms and warnings	5 3	6 0	6 2
5 4 4 9 Effectiveness of training system for training processor time in digital mode versus a real-time audio count rate	5 3	5 8	5 9
5 4 5 10 Effectiveness of training system for training the repair of the operational probe windows	4 2	4 5	4 6
5.5 3 Effectiveness of training system for simulating the contamination of the probe following contact with a contaminated surface	4 2	4 9	3 9

In Table 4 three questions (5.4.4.2, 5.4.4.3, and 5.4.4.7) were not tallied and indicate NA. The effectiveness of the training system for training wide airborne contamination by multiple gamma, alpha, or beta emitters (e.g. in a plume) (5.4.4.2) was not tallied because airborne contamination could not be simulated for alpha and beta (not permitted for safety reasons) and therefore was not reasonable to assess. The effectiveness of the training system for training changes in contamination according to user-defined winds or radioactive decay rates (5.4.4.3) could not be assessed as a result of an inability to connect the GPS for satellite computer gamma simulations and therefore it was not reasonable to assess. Finally, the effectiveness of the training system for training statistical fluctuations in the rate count (5.4.4.7) was not tallied because participants could not assess this technical requirement during the usability trial.

As the intent of the trial was to assess the usability of both operational and training systems, evaluation questions were designed and based upon the RFP. However, during the trial it was determined that specific questions, as detailed above, could not be fairly assessed. Therefore to ensure a fair evaluation for all systems eleven questions were removed and not reported on.

Discussion

Results from the questionnaires revealed that the overall combined weighted scores for questions retained totalled 584.3 for System 1 (70%), 682.0 for System 2 (81%) and 756.9 for System 3 (90%) out of a total possible score of 840. Participants indicated that System 3 was the most desired system out of the three systems for operations, set-up and maintenance, and training. This satisfaction was evident when participants were asked to select the best systems based upon display characteristics and lighting, 11 of the 12 voted for System 3 while one voted for System 2. For ease of maintenance all 12 participants voted for System 3, and when asked to choose the best system based upon operational and maintenance capability, System 3 received 11 votes and System 2 received one vote. The one vote for System 2 was based upon the perception that System 2 did not need a computer therefore reducing the complexity of set up and the likelihood of failure. When participants were requested to select the best training system, eleven of the participants voted for System 3 as the best training system while one subject abstained. Thus, when asked to choose the best overall system to replace that currently in service, it is not surprising that the participants unanimously voted for System 3.

Operations

Under the operations portion of the usability trial System 1's mean scores for indoor and outdoor operations were consistently between barely unacceptable and borderline. The exception being the durability of the storage case which rated reasonably acceptable (6.1), and ease of distinguishing the cases which rated barely acceptable (5.9). System 2's mean scores were marginally better than System 1, where twelve of the 15 components evaluated in the indoor and outdoor test stands were within the borderline to barely acceptable range. The remaining 3 criteria, which included: ease of operating the controls (6.1); durability of storage case (6.3); and ease of use of the storage case (6.0), rated reasonably acceptable. System 3's mean scores rated reasonably to completely acceptable in all criteria in both the indoor and outdoor test stands, with the exception of prevention of contact with contaminated surfaces which scored barely acceptable (5.3) as did effectiveness for protecting the probe (5.6). Participants gave System 3 these two barely acceptable ratings as a result of problems they had with the hanging probe cover.

During the focus group, System 1 was universally regarded as unacceptable while System 3 was universally accepted. System 2 was regarded as in between unacceptable and accepted. These results are generally consistent with the overall ratings from the questionnaire. Participants discussed the areas of concern with each system, and found System 1 to be heavy and awkward, indicating the system was very difficult to operate with the extension probe while simultaneously looking at the meter and controlling the probe extension. These responses validate the questionnaire rating associated with preventing inadvertent contact with contaminated surfaces, ease of using the probe extension and using the system to conduct survey (scoring 3.2, 3.9 and 3.5 respectively). Participants found that the probe had to be removed from the source to read the meter. As well, the wand extension was not durable and bent easily under the weight of equipment. System 2 did fare better on weight and ease of use issues (4.0, 4.9 and 4.3, respectively). However, comments in the focus group indicate the system was also cumbersome and awkward to use, and without handles the distribution of weight was uneven. System 3 scored best on the ease of use issues (5.3, 6.8 and 6.4, respectively). Focus group discussions indicated that of the three systems, System 3 was found to be an excellent system to use, due to the weight, balance and type of extension used.

Questionnaire scores associated with compatibility aligned with focus group compatibility discussions for the three systems. System 1 scored low on compatibility of system components with combat clothing and compatibility of system components with NBC TOPP high (4.5, 3.4). Participants reported they had a hard time removing the headset connection from System 1's meter with their bare hands and when wearing NBC gloves. Participants found the closeness of the buttons on System 1 made it difficult to operate and handle, and with many small pieces they believed the risk of losing pieces was high. System 2 scored better than System 1 on compatibility of system components with combat clothing and compatibility of system components with NBC TOPP high (4.8, 4.4). However, when surveying with the system slung on the chest, as recommended, visual obstructions made the meter difficult to read. Participants found System 2's audible alarm hard to hear in presence of background noise when wearing NBC gear. System 3 scored well on compatibility of system components with combat clothing and compatibility of system components with NBC TOPP high (6.5, and 6.1). Participants found the system easy to assemble and handle, easy to survey and not impeded by masks and gloves.

As to the IRDMS's durability, System 1 (4.8) scored the same durable rating as System 2 (4.8) while System 3 scored a higher rating to both System 1 and 2 (6.3). Discussions revealed System 1 was cumbersome, parts had to be placed exactly in the case or there was a chance of damage upon closing. While with System 2, participants found the alpha Mylar sensory face was very easy to damage when scanning grass. However, when the Mylar sensory face was damaged during the trial the liquid repair was found to be very effective. As well, even though System 2's cases were durable (rating 6.3), the two cases were difficult to carry around and subjects indicated it would be more practical to fit all components into one case. System 3 was not found to have significant issues regarding durability. The only important issue participants raised was that the gamma/beta probe cover was susceptible to damage and contamination while performing a survey. They rated this problem 5.6, suggesting that if possible the cover should be attached to the probe when in use. Participants did not consider this a serious enough issue during use to deem it an unacceptable quality, in fact they suggested a Velcro strap would easily resolve the problem.

In terms of system assembly, participants found System 1 probes could be easily changed; however, they found there were too many probes. Perhaps System 1's alpha and beta probes should be combined into a single probe. With System 2 they found it was difficult to connect the probes to the meter, and with System 3 the set up of the alpha probe was found to be too complex by participants.

When operating controls, System 2 scored almost as well as System 3 on ease of operation (6.1 for System 2 and 6.3 for System 3) while System 1 obtained a borderline rating (4.1). Focus group comments indicate System 2's control buttons were very easy to use, the fast refresh was good and it was easy to determine which alpha/beta source was present with a flick of a button. No apparent control operating issues were raised for System 3 in the focus group discussions. The focus group stated the System 1 was complex and difficult to use.

In the focus group participants made it clear that operationally they preferred System 3. When participants were asked to choose between the systems for the best operational system the results were System 1 – 0, System 2 – 0, System 3 – 12.

Set-up and Maintenance

In the Set-up and Maintenance evaluation System 1's mean scores lay between barely acceptable and reasonably acceptable. The exceptions were readability of visual displays in low light/night conditions (4.4) and ease of setting the alarm set points (4.8). For readability of visual display in bright sunlight conditions System 1 scored the highest (6.4). System 2 and System 3's scores ranged between barely acceptable to reasonably acceptable. Focus group discussions indicate that System 1 was universally regarded as unacceptable, while Systems 2 and 3 were universally accepted, with System 3 being favoured over System 2. These results are not entirely consistent with the overall ratings from the questionnaire. This inconsistency could not be discerned in the discussions but may be due to the participants not considering some of the issues related to the systems until they were discussed at the focus group. It may also be due to the influence of other's opinions during the focus group.

Results also show that System 3 was superior to System 1 and slightly superior to System 2 in display readability and ease of operation. System 1 scored low on readability of visual displays in low light/night conditions (4.4) and focus group statements reinforce this score. Comments indicate the display was dark, used very small numeric values, and did not use a bar display, all of which made readability difficult in night conditions. System 2 scored better on readability (6.6) but this contrasted with focus group discussions where it was stated that the display of System 2 was too small, the contrast was not sharp and the visual display should be bigger. Even though System 3 scored high on readability (6.9), the focus group stated that the display of System 3 was hard to see in the glare of the sun.

It was also found that System 2 was superior to both System 1 and System 3 for setting alarms. System 1 scored lower than System 2 on all questions relating to ease of use and effectiveness of alarms (4.8 to 5.8 for System 1 vs. 5.7 to 6.5 for System 2). These scores aligned with focus group comments. Participants stated that with System 1 there was a noticeable delay between radioactive source contact and the alarm's activation (about 1 or 2 seconds). System 1 was found to have a short time lapse before the alarm cleared, thus, if the alarm was the only detection method, a source could be missed, or the system would detect contamination when none was present. System 1 had no alarm volume control and no auditory low battery warning. Participants also found they required a certain proficiency to set Systems 1's alarms and dose rates, the controls were complicated and could be easily pushed by mistake. For example, three buttons must be pushed to reset dose level to zero, and the alarm threshold setting requires the unit to first be switched off prior to resetting. In regards to System 2, focus group discussion revealed that even though a complex operating process must be learned to set alarms, the alarm worked well once the process was mastered (6.5). Participants found there was a great advantage to having one probe for alpha/beta detection with a built in library. They also found the auditory alarms for System 2 were strong and easily heard (6.5)however, participants indicated inclusion of a visual light alarm would make the system more effective. While discussing System 3 in the focus group, participants stated the dose and dose rate alarm thresholds could only be changed with a personal computer system, and participants found this to be a disadvantage and cumbersome. Participants indicated it would be a difficult task to perform this in the field, and compounding this, the risk of human error increases when another source is used for calibration. As well, participants indicated System 3's alarm was found to be difficult to hear.

Although all systems were found to have deficiencies, when asked to choose between the systems for Set-up and Maintenance participants voted System 1 – 0, System 2 – 1, System 3 – 11. The one person who voted for System 2 did so because of the perception that System 3 was computer dependent increasing the learning curve and chance of making a mistake, while System 2 was considered simple and easy.

Training

In the Training System evaluation, System 1's training mean scores ranged from borderline to barely acceptable. The exception was the question dealing with effectiveness of training system for training widely dispersed ground contamination by multiple gamma, alpha or beta emitters. This received a barely unacceptable rating (3.3), and participants stated that System 1 was not effective for dealing with all three types of radiation equally. System 2's mean

scores were better than System 1's mean scores, ranging from barely acceptable to reasonably acceptable; however, there were exceptions. Questions dealing with effectiveness of training system for training the repair of the operational probe window and effectiveness of training system for simulating the contamination of the probe following contact with a contaminated surface were found to be borderline (4.5 and 4.9 respectively) for System 2. Participants found that they could easily damage the probe. System 3's scores ranged from borderline to reasonably acceptable, with the exception of the question regarding effectiveness of the training system for simulating the contamination of the probe following contact with a contaminated surface, which rated barely unacceptable (3.9) because the participants found that probe cover was easily contaminated.

In the Training system focus group discussions System 1 was universally regarded as unacceptable, while System 2 and System 3 were regarded as universally accepted. These results were consistent with the overall ratings from the questionnaires, where System 2 and System 3 were both rated as acceptable. System 2's combined training scores did achieve a slightly higher mean acceptability rating than System 3 (System 2, 5.5 vs. System 3, 5.4). This slight difference may be due to System 3's training prototype not representing the operational probe, therefore, influencing the overall opinions of the participants in the suitability of System 3.

System 1 did poorly on questions relating to effectiveness to detect contamination as shown by the relevant scores being borderline or below (4 or less). Although the training alpha and beta probes were lightweight and reasonably easy to use, the probes were triggered by any metallic source, resulting in many false positives when using the probe within or near any metal surface. Focus group discussions revealed that this could be considered a significant shortfall of the system as conducting training within a metal environment could be difficult to almost impossible. A false positive problem with System 2 was also discussed in the focus group, participants reported that System 2 alpha-beta training probe was overly sensitive and reacted where no compound was placed. As well, participants found System 2's blue light detection system required it to be directly over contamination before obtaining a reaction. The previous two statements could seem a contradiction of System 2's detection capability, indicating both a false positive and false negative tendency. This discrepancy can be explained by the fact that training sources were the cause of false positive readings, while the false negative was a result of the small detection portion in System 2's probe. Participants did rate System 3 lower than System 2. Even though System 3 scored less than System 2, focus group discussions revealed that System 3 was a good training system. The focus group stated that System 3 was reliable, the beta and alpha probes gave excellent warning and the agents were easy to use. This may appear to be a discrepancy, however, the design of System 3's training beta probe did not represent the operational probe and participants felt it needed to be changed to reflect the operational system. Participants found System 3 to be very sensitive to the magnetic and chemical check sources providing realistic readings, a very small concentration of liquid elicited a reaction that simulated contamination. Even though System 3's alpha probe worked very well with the training check sources the probe could not be contaminated when near, or in, the radioactive training sources. For this reason participants gave a low score for both effectiveness of training system for simulating the contamination of the probe following contact with a contaminated surface (3.9) and effectiveness of training system for training probe contamination when contacting the contaminated surface (4.3). In

order for the training system to be realistic there must be a possibility of contaminating the training probe.

When looking at the gamma training portion, System 1's GPS training system worked well on the computer, however, System 2's GPS training system required all simulation information to be stored in each probe via a card prior to use, making the system much more cumbersome and less adaptable for field use. The participants found System 3's GPS training system to work well. It was suggested however, that System 3's GPS system should be housed in the gamma meter vice a unique GPS device.

It is interesting to note that for the question ease of using the training system, all 3 systems received approximately the same score (System 1, 5.6, System 2, 5.7, System 3, 5.6) whereas ease of learning the controls the scores varied (System 1, 4.1, System 2, 6.1, System 3, 6.3). This discrepancy may reflect prior learning from operating the operational system earlier in the study.

Focus group discussions raised several questions about the training system concept that could not be answered at the time and may need to be investigated further. For instance, will the training simulation chemicals work in wet and snowy conditions, and will the training simulation chemicals affect the mechanisms of any device placed on the meter?

When participants were asked to choose between the systems for the best Training System they voted System 1 – 0, System 2 – 0, System 3 – 12.

Ratings and Summary

In the focus groups various characteristics of the three systems were compared and voted on. When asked to choose between the three Systems based upon display characteristics; lighting; durability of meter, probe, and extension, how effective the probe worked; and ease of assembling the system, the results indicated System 1 – 0, System 2 – 1, System 3 – 11. The voting results between the three systems based upon alarm set-up, alarm indicator and battery maintenance were System 1 – 0, System 2 – 4, System 3 – 7. Choosing between the three systems based upon ease of maintenance and system sensitivity participants voted System 1 – 0, System 2 – 0, System 3 – 12. Participants listed the three systems in order of sensitivity, System 3 was the most sensitive, System 2 was considered to be less sensitive than System 3 and System 1 was considered the least sensitive. When asked to make an overall rating of the three systems based upon set-up maintenance and ease of use, the results were System 1 – 0, System 2 – 1, System 3 – 11. The one person who voted for System 2 did so because of a perception that a computer was required to set System 3 up which made the system computer dependent. Participants were then asked to make an overall rating for the best training system unanimously voting for System 3 (12).

The questionnaire results of the study coincided strongly with focus group discussions on which system participants rated the highest. For operational aspects System 3 was found to be the best system. For set-up and maintenance System 3 was favoured but System 2 was liked as well. In regards to the training system, System 3 was favoured, as was System 2. Although System 3 was favoured, this choice may be based on its operational capability as System 3's training system was seen to have its faults.

Conclusions

Compliance with Request for Proposal and Improvements

All three of the systems obtained borderline ratings and two of the systems obtained unacceptable ratings that may conflict with their compliance with the RFP. To add value to the assessment of a successful selection participants were requested to provide recommendations for all systems so that if awarded there would be a list of specific improvements. As this usability evaluation was only a portion of the IRDMS evaluation to select the best system, the instrument that received the highest usability score (System 3) may not in fact be the winning bid.

The focus group did agree that overall any IRDMS adopted must have various basic features: alarm volume controls; auditory low battery warning; and headphones. In addition, the focus group agreed that any IRDMS training system must have a map display with a clock. Participants stated that the clock should have the capability to track the change in radiation conditions over time and show bomb yield strength.

Given the unacceptable ratings participants felt that if System 1 were to win the bid evaluation the following improvements were suggested as mandatory. The alpha and beta probes must be combined into a single probe, since switching between probes while performing a survey was not efficient and could be difficult in a field environment. Due to the potential risk of human error it was suggested that the toggle switches be replaced with a type of button that could not be accidentally switched. Also, it was suggested that the wand extension be more durable so that it does not bend and result in contamination. Due to the extension bending participants had to hold the extension in an awkward position increasing arm and back strain thus increasing the length of survey time as frequent operator breaks were needed. A non-metal method of detection for the training system must be developed so that the system can be used in all operational environments, to include onboard ships or near vehicles.

System 2 received no unacceptable ratings but did have fourteen borderline ratings. Given this, System 2 was rated as second best, however, participants felt that if System 2 were to win the bid evaluation the following suggestions were made should this system be procured. The system needs to incorporate a strap assembly so that the system can be used hands free as the beta probe with the wand extension could not be used for 20 minutes comfortably. The system needs to be lighter and the extension wand must be reinforced so it does not twist. The carrying/storage cases must be reduced from two to one case for carrying ease. The training system's training probes should have more light detectors on them to make the training probe as sensitive as the operational probe. This was deemed as a critical improvement as the training probes need to represent the operational probes in order to simulate a realistic operational environment for training purposes.

Although System 3 was rated as the best of the three systems, it did receive one unacceptable rating and two borderline ratings, and participants did suggest some improvements if this system was to be procured. It was felt that a tilt lever and mounting strap on the probe

extension would reduce operator back and arm strain. The meter housing was deemed to be flimsy, and it was suggested it be made more durable for field use. For the training system it was also suggested that the training probes should have more or larger detector fields to allow the training probe to reflective of the operational probe.

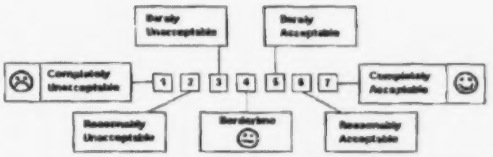
References

1. Request for Proposal W8476-000187/A (IRDMS RFP).
2. IRDMS Usability Evaluation Plan 25 May 2001.

Annex A
IRDMS Usability Evaluation Questionnaires

NAME: _____

IRDMS: _____



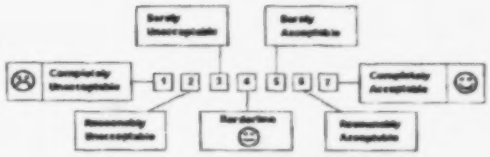
Rate the following features using the 7-point acceptability scale above.	**IRDMS**
OPERATION	1 2 3 4 5 6 7
3.3.1 Ease of operating the controls.	$ $ $ $ $ $ $
3.6.8 a) Compactness of the operational probes.	$ $ $ $ $ $ $
3.6.8 b) Weight of the operational probes.	$ $ $ $ $ $ $
3.6.12 Effectiveness of the system for preventing inadvertent contact with contaminated surfaces.	$ $ $ $ $ $ $
3 7.4 Ease of using the probe extension.	$ $ $ $ $ $ $
3.9.2 Ease of using the system to survey for up to six hours on a 20 minute work / 10 minute rest cycle.	$ $ $ $ $ $ $
3.9.1 b) Compactness of system (i e. meters, cases, probes, and extension wands).	$ $ $ $ $ $ $
3.9.1 c) Weight of system (i.e. meters, cases, probes, and extension wands)	$ $ $ $ $ $ $
3.9.3 a) Compatibility of System 3omponents with **combat clothing** (i e. combats, combat gloves, and sand/wind/dust goggles)	$ $ $ $ $ $ $
3.9 3 b) Compatibility of System 3omponents with **NBC TOPP high** (i.e NBC IPE, NBC gloves, overboots, and C4 mask).	$ $ $ $ $ $ $
DURABILITY	1 2 3 4 5 6 7
3.6.11 Effectiveness of mechanism for protecting the probe, when in use.	$ $ $ $ $ $ $
3.8.1 a) Durability of storage cases.	$ $ $ $ $ $ $
3 8.1 b) Ease of using storage cases.	$ $ $ $ $ $ $

3.8.5 Ease of distinguishing storage cases.	$ $ $ $ $ $ $
3.9 1 a) Durability of system (i e. meters, cases, probes, and extension wands).	$ $ $ $ $ $ $
TRAINING SYSTEM Vs OPERATIONAL SYSTEM	**1 2 3 4 5 6 7**
5.4.2 Similarity of information displayed by the training system to the operational system.	$ $ $ $ $ $ $
5.4.2 Similarity of operation of the training system to the operational system	$ $ $ $ $ $ $

COMMENTS: Please use the back of this sheet for any additional comments

NAME: _____

IRDMS: _____

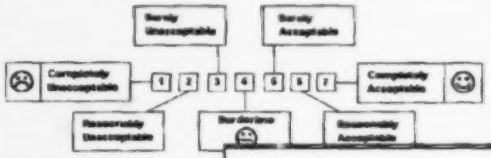


Rate the following features using the 7-point acceptability scale above.	IRDMS
CONFIRM OPERATION	1 2 3 4 5 6 7
3 2.6 a) Ease of comprehending the **visual** display changes in radioactivity levels.	$ $ $ $ $ $ $
3.2.6 b) Ease of comprehending the **auditory** display changes in radioactivity levels.	$ $ $ $ $ $ $
3 2.7 a) Readability of visual display in **normal daylight** conditions.	$ $ $ $ $ $ $
3.2.7 b) Readability of visual display in **bright sunlight** conditions.	$ $ $ $ $ $ $
3.2.7 c) Readability of visual display in **fluorescent lighting** conditions.	$ $ $ $ $ $ $
3.2 7 d) Readability of visual display in **low light / night** conditions.	$ $ $ $ $ $ $
3.2.8 a) Ease of silencing the audible display.	$ $ $ $ $ $ $
3.2 8 b) Speed of silencing the audible display	$ $ $ $ $ $ $
3 2.9 Ease of operating the volume control.	$ $ $ $ $ $ $
3.2.10 Effectiveness of the earphones in high noise conditions.	$ $ $ $ $ $ $
3.2.11 Ease of subtracting background radiation levels during alpha or beta surveys.	$ $ $ $ $ $ $
3.3 2 Ease of switching between gamma, alpha, and beta radiation readings.	$ $ $ $ $ $ $
SET ALARMS	1 2 3 4 5 6 7
3.4 2 a) Effectiveness of the **visual** alarm.	$ $ $ $ $ $ $
3 4.2 b) Effectiveness of the **auditory** alarm.	$ $ $ $ $ $ $
3.4.5 a) Ease of deactivating/silencing audible alarm **before** alarm level is reached.	$ $ $ $ $ $ $

3.4.5 b) Ease of deactivating/silencing audible alarm **after** alarm level is reached.	$ $ $ $ $ $ $
SET ALARMS (continued)	**1 2 3 4 5 6 7**
3.4.6 Ease of setting the alarm set points.	$ $ $ $ $ $ $
3.4.7 a) Ease of activating the alarms.	$ $ $ $ $ $ $
3.4 7 b) Ease of deactivating the alarms.	$ $ $ $ $ $ $
MAINTENANCE	**1 2 3 4 5 6 7**
3.10.3 Ease of replacing batteries in the field.	$ $ $ $ $ $ $
3.10.7 a) Effectiveness of **auditory** "low battery" warning.	$ $ $ $ $ $ $
3.10.7 b) Effectiveness of **visual** "low battery" warning.	$ $ $ $ $ $ $
COMMENTS: Please use the back of this sheet for any additional comments	

NAME: _____

IRDMS: _____



Rate the following features using the 7-point acceptability scale above.	**IRDMS**
CONTROLS AND PROBES	**1 2 3 4 5 6 7**
3.3 1 a) Ease of learning the controls.	$ $ $ $ $ $ $
3.3.1 b) Ease of comprehending the control markings.	$ $ $ $ $ $ $
3.6.8 a) Compactness of the training probes.	$ $ $ $ $ $ $
3.6 8 b) Weight of the training probes.	$ $ $ $ $ $ $
TRAINING SYSTEM	**1 2 3 4 5 6 7**
5.2.3 Ease of using the training system.	$ $ $ $ $ $ $
5.4.1 Durability of the training system.	$ $ $ $ $ $ $
5.4.4.1 Effectiveness of training system for training **widely dispersed ground contamination by multiple gamma, alpha, or beta emitters.**	$ $ $ $ $ $ $
5.4 4 2 Effectiveness of training system for training **wide airborne contamination by by multiple gamma, alpha, or beta emitters (e.g. in a plume).**	$ $ $ $ $ $ $
5.4.4.3 Effectiveness of training system for training **changes in contamination according to user-defined winds or radioactive decay rates.**	$ $ $ $ $ $ $
5.4.4.4 Effectiveness of training system for training **the detection of agents only within a few centimetres of the contamination surface.**	$ $ $ $ $ $ $
5.4.4 5 Effectiveness of training system for training **probe contamination when contacting the contaminated surface.**	$ $ $ $ $ $ $
5 4 4.6 Effectiveness of training system for training **monitoring during and after decontamination of personnel and equipment.**	$ $ $ $ $ $ $
5.4.4 7 Effectiveness of training system for training **statistical fluctuations in the rate count.**	$ $ $ $ $ $ $
5.4 4.8 Effectiveness of training system for training **alarms and warnings.**	$ $ $ $ $ $ $

TRAINING SYSTEM (continued)	1	2	3	4	5	6	7
5 4.4.9 Effectiveness of training system for training **processor time in digital mode versus a real-time audio count rate.**	$	$	$	$	$	$	$
5.4.5.10 Effectiveness of training system for training **the repair of the operational probe windows**	$	$	$	$	$	$	$
5.5.3 Effectiveness of training system for simulating the contamination of the probe following contact with a contaminated surface.	$	$	$	$	$	$	$

COMMENTS: Please use the back of this sheet for any additional comments

List of Abbreviations

DND	Department of National Defence
GPS	Global Positioning System
IRDMS	Ionizing Radiation Detection Meter Systems
Maint	Maintenance
NBC	Nuclear, Biological and Chemical
NCM	Non Commissioned Member
Ops	Operational
RadSO	Radiation Safety Officers
RFSO	Request for Standing Offer
RFP	Request for Proposal
Sv	Sievert
TOPP	Theatre Orientation Protective Posture
Trg	Training
μSv	MicroSievert

DOCUMENT CONTROL DATA SHEET

1a PERFORMING AGENCY	2 SECURITY CLASSIFICATION
DRDC Toronto	UNCLASSIFIED Unlimited distribution -
1b PUBLISHING AGENCY DRDC Toronto	

3 TITLE

(U) Ionizing Radiation Detection Meter Systems usability evaluation

4 AUTHORS

Yvonne Severs Robert Hughes David Tack Linda L. M. Bossi

5. DATE OF PUBLICATION	6. NO OF PAGES
November 1 , 2002	41

7 DESCRIPTIVE NOTES

8 SPONSORING/MONITORING/CONTRACTING/TASKING AGENCY

Sponsoring Agency
Monitoring Agency
Contracting Agency .
Tasking Agency Canadian Forces Nuclear Detection, Identification and Dosimetry project Connaught Ranges, Ottawa

9. ORIGINATORS DOCUMENT NO	10 CONTRACT GRANT AND/OR PROJECT NO	11 OTHER DOCUMENT NOS
Technical Memorandum TM 2002-208		

12 DOCUMENT RELEASABILITY

Unlimited distribution

13 DOCUMENT ANNOUNCEMENT

Unlimited announcement

14 ABSTRACT

(U) In support of the Canadian Forces Nuclear Detection, Identification and Dosimetry project a usability trial of three Ionizing Radiation Detection Meter Systems (IRDMS) was conducted 18-22 June 2001 at Connaught Ranges. The systems evaluated: System 1: ANPDR 77 Operational System with PDR 77 Alpha/Beta probes and GPS (Global Positioning System) Training System; System 2: ADM 300C Operational System with ABP100 Alpha/Beta probes and GPS Training System; and System 3: RAM R200 Operational System with PA 100M and RG12 Alpha/Beta probes and GPS Training System. Twelve military personnel (Private to Captain) determined compliance with the bid evaluation criteria. The counter-balanced study, based on a Latin square design, was divided into two assessment phases, an operational phase and a training system phase, which rotated three groups (of four participants each) concurrently through each IRDMS. Groups were trained in set-up, maintenance and operation of each bid IRDMS prior to evaluation. The operational evaluation consisted of a test/maintenance stand, an indoor/outdoor operational stand, and an indoor/outdoor training system stand. Participants rated the utility and usability of each system, by measuring ease of use/maintenance under a range of relevant operational and clothing conditions, compatibility with soldier tasks, compatibility with clothing, compatibility with equipment and operational environments, and ease of learning (including an assessment of initial and refresher training implications). A 7-point Likert-like acceptability rating scale, which covered a range from completely unacceptable (1) to completely acceptable (7), was used to evaluate the measures. The combined weighted mean scores were 584.3 for System 1 (70%), 682.0 for System 2 (81%) and 756.9 for System 3 (90%) out of a total possible score of 840. Participants reported mean scores for System 1 between barely unacceptable and borderline for both indoor and outdoor operations. System 2 was better, whereby 11 of the 15 questions resulted in mean scores within the borderline to barely acceptable range. System 3 rated reasonably to completely acceptable on all questions, with the exception of prevention of contact with contaminated surfaces which scored barely acceptable. Under set-up and maintenance criteria little difference was found between the 3 systems, although System 1 rated consistently lower than System 2 which rated consistently lower than System 3. System 1 consistently rated barely acceptable to reasonably acceptable, and System 2 barely acceptable to completely acceptable. System 3 was also consistent, whereby 15 of the 16 criteria ranged between reasonably acceptable to completely acceptable. In the training system ratings, System 1 ranged from borderline to barely acceptable, System 2 ranged from barely acceptable to reasonably acceptable and System 3 ranged from reasonably acceptable to completely acceptable. Focus group discussions indicated that System 1, with an extremely small display and function buttons was too complex, with the potential for creating performance error. Another key problem was that participants could not comfortably operate System 1 and System 2 with the extension probe for 20 minutes, a realistic requirement in an operational setting. System 3 was found to be comfortable and balanced. Although some minor changes were suggested in focus group discussions System 3 was found to be the most acceptable system operationally, with an acceptable training system.

(U) En soutien au projet de détection, d'identification et de dosimétrie nucléaires des Forces canadiennes, nous avons effectué au polygone de Connaught, entre les 18 et 22 juin 2001, des essais sur la facilité d'emploi de trois systèmes de détection et de mesure de la radioactivité. Nous avons évalué les trois systèmes suivants : (1) un appareil opérationnel AN?PDR 77 avec des sondes alpha/bêta PDR 77 et un système GPS (Système mondial de localisation) pour l'entraînement; (2) un appareil opérationnel ADM 300C avec des sondes alpha/bêta ABP100 et un système GPS pour l'entraînement; (3) un appareil opérationnel RAM R200 avec des sondes alpha/bêta PA 100M et RG12 et un système GPS pour l'entraînement. Douze membres des forces (des grades de soldat à capitaine) ont mesuré la conformité des systèmes avec les critères d'évaluation des soumissions. L'étude équilibrée, basée sur le carré latin, comportait deux phases d'évaluation : une phase opérationnelle et une autre d'entraînement à l'utilisation du système. Trois groupes de quatre participants ont utilisé chaque système en succession. On a formé les groupes au réglage, à l'entretien et à l'utilisation de chaque système avant qu'ils les évaluent. L'évaluation opérationnelle comportait un banc d'essai et d'entretien, un autre pour évaluer le fonctionnement à l'intérieur et à l'extérieur et un dernier banc pour la formation à l'intérieur ou à l'extérieur. Les participants ont évalué l'utilité et la facilité d'emploi de chaque système en évaluant s'ils étaient faciles à utiliser et à entretenir pour une gamme de conditions opérationnelles et de tenues militaires; sa compatibilité avec les tâches des soldats, la tenue portée, l'équipement et l'environnement des opérations; ainsi que la facilité d'apprentissage (notamment l'évaluation de l'effet des formations

initiales et d'appoint). Pour évaluer les mesures, nous avons utilisé une échelle de notation de sept points, analogue à celle de Likert, s'étendant de « complètement inacceptable » (note 1) à « complètement acceptable » (note 7). Nous avons obtenu des cotes combinées des moyennes pondérées de 584,3/840 (70 %) pour le système 1, de 682,0/840 (81 %) pour le système 2 et de 756,9/840 (90 %) pour le système 3. Les participants ont attribué au système 1 des notes moyennes entre « à peine inacceptable » et « médiocre » pour son utilisation à l'intérieur ou à l'extérieur. Ils ont mieux aimé le système 2 et lui ont attribué une note moyenne allant de « médiocre » à « à peine acceptable », à 11 questions sur 15. Les participants ont accordé au système 3 des notes allant de « raisonnablement acceptable » à « complètement acceptable » pour toutes les questions, hormis celle sur la prévention du contact avec les surfaces contaminées à laquelle ils ont répondu : « à peine acceptable ». Les différences entre les trois systèmes quant aux critères du réglage étaient minimes. Cependant, les notes du système 1 étaient toutes plus basses que celles du système 2, lesquelles étaient toutes plus basses que celles du système 3. Les participants ont généralement attribué au système 1 une note allant de « à peine acceptable » à « raisonnablement acceptable » et au système 2, de « à peine acceptable » à « complètement acceptable ». Les notes attribuées au système 3 étaient cohérentes : à 15 questions sur les 16, les participants ont attribué une note allant de « raisonnablement satisfaisant » à « complètement satisfaisant ». Au chapitre des notes sur l'entraînement, on a donné au système 1 une note allant de « médiocre » à « à peine acceptable », au système 2 de « à peine acceptable » à « raisonnablement acceptable » et au système 3 de « raisonnablement acceptable » à « complètement acceptable ». Le groupe de discussion a indiqué que l'affichage trop petit et les boutons de fonction trop complexes du système 1 pouvaient provoquer des erreurs de fonctionnement. Les essais ont révélé un autre problème important : les participants ont indiqué qu'ils ne pouvaient pas utiliser aisément les systèmes 1 et 2 avec la sonde pendant 20 minutes, une exigence prévisible dans un contexte opérationnel. On a trouvé que le système 3 était d'un emploi facile et était mécaniquement équilibré. Le groupe de discussion a suggéré des modifications mineures au système 3, lequel a été considéré comme le système le plus acceptable pour les opérations, après un entraînement adéquat.

15 KEYWORDS, DESCRIPTORS or IDENTIFIERS

(U) Ionizing Radiation; IR instrument detection; IR detection; monitoring systems; IRDMS

Defence R&D Canada

Canada's leader in defence
and national security R&D

R & D pour la défense Canada

Chef de file au Canada en R & D
pour la défense et la sécurité nationale

#518903
CA022091

DEFENCE DÉFENSE

www.drdc-rddc.gc.ca